The effects of shared storybook reading on word learning: a meta-analysis

Article (Accepted Version)

The Effects of Shared Storybook Reading on Word Learning: A Meta-Analysis

Zoe M. Flack, Andy P. Field & Jessica S. Horst*

University of Sussex

Keywords: word learning, vocabulary, shared storybook reading, meta-analysis, dialogic reading style

Acknowledgements:

We thank Rod Bond and Ellen Thompson for helpful comments on the analysis. We would also like to thank Jeanne Shinskey and Amy Booth for their assistance with unpublished data.

*Dr. Jessica S. Horst
jessica@sussex.ac.uk
University of Sussex
School of Psychology
Pevensey 1 Building
Brighton, BN1 9QH, UK
+44 (0)1273 873084

Ms. Zoe M. Flack
z.flack@sussex.ac.uk
University of Sussex
School of Psychology
Pevensey 1 Building
Brighton, BN1 9QH, UK
+44 (0)1273 876561

Prof. Andy P. Field
Andy.field@sussex.ac.uk
University of Sussex
School of Psychology
Pevensey 1 Building
Brighton, BN1 9QH, UK
+44 (0)1273 877150

Abstract

Although a rich literature documents pre-literate children's word learning success from shared storybook reading, a full synthesis of the factors which moderate these word learning effects has been largely neglected. This meta-analysis included 38 studies with 2,455 children, reflecting 110 effect sizes, investigating how reading styles, story repetitions, tokens and related factors moderate children's word comprehension, while adjusting for the number of target words. Dialogic reading styles, tokens, and the number of words tested all moderated word learning effects. Children's age, who read the story, and time between story and test were not moderators. We identify story repetition and word types as topics which merit further research. These results provide information to guide researchers and educators alike to the factors with the greatest impact on improving word learning from shared storybook reading.

Keywords: word learning, vocabulary, shared storybook reading, meta-analysis, dialogic reading style

The Effects of Shared Storybook Reading on Word Learning: A Meta-Analysis

Shared storybook reading provides several benefits to young children including parent-child bonding (Barratt-Pugh & Rohl, 2015; Schwartz, 2004), fostering a love of reading later in life (Bus, 2001; Pillinger & Wood, 2014) and learning to sustain attention (Lawson, 2012). Much of children's developing lexicon is encountered through everyday conversation (Weizman & Snow, 2001), but shared storybook reading provides a complementary source of vocabulary (Montag, Jones, & Smith, 2015). Because vocabulary size at school entry predicts later academic achievement (Coyne, Simmons, Kame'enui, & Stoolmiller, 2004; Sénéchal, LeFevre, Thomas, & Daley, 1998), understanding how to help children maximize word learning from shared storybook reading can provide feasible interventions for education. For the purposes of this paper, we focused on word comprehension (the understanding of what a word refers to) rather than word production (the ability to use a word verbally at any moment), although word learning may refer to either comprehension or production in the shared storybook reading literature. Numerous investigators have demonstrated that shared storybook reading promotes word comprehension (e.g., Justice, Meier, & Walpole, 2005; Sénéchal & Cornell, 1993); however, a clearer understanding of the strength of these effects is wanting.

Several factors are reported to influence word learning from storybooks, including who reads (Hindman, Connor, Jewkes, & Morrison, 2008), how they read (Bus, van Ijzendoorn, & Pellegrini, 1995; Reese & Cox, 1999), the child's age (Hargrave & Sénéchal, 2000), the types of words being taught (e.g., McLeod & McDade, 2011; Storkel & Maekawa, 2005), the number of words in the story (e.g., Robbins & Ehri, 1994), and how many times the story is heard (e.g., Horst, Parsons, & Bryan, 2011; Sénéchal, 1997). For example, Whitehurst & colleagues (1988) found more interactive reading styles positively influence word learning from storybooks. Specifically, they taught parents to use dialogic styles such as open-ended questions, pointing, repetition and generally encouraging text-related talk during reading. The benefits of such dialogic styles have been widely reported (e.g., Biemiller & Boote, 2006; Justice et al., 2005; Sénéchal, 1997). Dialogic styles represent relatively simple interventions, which can be easily taught to parents or teachers to enhance word learning.

Although mothers' reading styles have garnered particular attention (e.g., Ninio, 1980; Reese & Cox, 1999), research examining the reading styles of both parents suggests any differences between mothers and fathers reading styles are rather subtle (Blake, Macdonald, Bayrami, Agosta, & Milian, 2006; Schwartz, 2004). This informs our understanding of how parents read with their children, but not how this influences subsequent word learning. Further, many of the experiments concerned with word learning from shared storybook readings are designed with experimenters or teachers as the reader, rather than parents. If the reader affects word learning, this would have important, practical implications for the generalizability of word learning research to naturalistic settings. In addition, a deeper understanding of the impact of different readers may inform future experimental designs (see also Aram & Besser, 2009).

Young children can learn up to four words each day (Bion, Borovsky, & Fernald, 2013; Fenson et al., 1994). More specifically, Biemiller and Boote (2006) compared 11 studies on shared storybook reading and reported word learning gains of approximately two words per day for 3- to 12-year-old children. Historically, children's ability to learn the meanings of new words was thought to be as high as up to nine new words per day (Carey, 1978), based on the observation of children fast mapping (quickly guessing the meaning of a new word based on lexical, syntactic and contextual cues). Although children may recognize a word after a single exposure under certain conditions (see Horst & Samuelson, 2008 for a review), robust learning the meanings of words requires multiple, repeated exposures (McMurray, Horst, & Samuelson, 2012). Increasing target word occurrences (tokens) within storybooks can provide these additional exposures (Elley, 1989). Reading the same story repeatedly can also provide multiple exposures. Several studies have examined the effect of different numbers of storybook repetitions on word learning as measured by forced-choice test trials of comprehension (e.g., Biemiller & Boote, 2006; Horst et al., 2011; McLeod & McDade, 2011; Sénéchal, 1997; Wilkinson & Houston-Price, 2013), but as yet, the true strength of this effect has not been examined.

Our literature review uncovered a wide range of approaches to investigating word learning from shared storybook reading. The number of target words within a story, the number of tokens, the kinds of words (e.g., nouns, verbs), the number of times a story is read, as well as who is reading, all vary greatly across studies. Similarly, how word learning is assessed also varies. The National Reading Panel (2000) suggests that researcher-developed assessments are more sensitive to word learning gains than existing standardized educational assessments that measure increases in general vocabulary. In practice, some experimental assessments measure growth in the total number of words children comprehend (i.e., general vocabulary), while others measure recall or production of specific target words from the text. Studies that include assessments of general vocabulary routinely involve interventions and longer timescales, (typically several weeks or more, see e.g., Aram & Besser, 2009; Reese & Cox, 1999; Valdez-Menchaca & Whitehurst, 1992). The dearth of meta-analyses addressing word learning from shared reading to date may well be explained by the high level of variability between measurement approaches. Although Marulis and Neuman (2010) investigated the effects of vocabulary interventions beyond shared storybook reading and Mol, Bus, de Jong, and Smeets (2008) investigated the effect of dialogic reading by parents, there are no meta-analyses exploring the effects of shared storybook reading on word learning in such breadth, despite a very rich contribution of research. However, school children may be able to learn up to 15% of target words when reading to themselves (see Swanborn & de Glopper, 2002, for a meta analysis).

Current study

The goal of the current meta-analysis was to estimate the population effect of shared storybook reading on word comprehension. We focus on word comprehension because it precedes word production (e.g., Huttenlocher, 1974). Despite the high level of variability noted above, studies of word comprehension are overall more similar in methods than studies of word production. Thus, by restricting our analysis to studies of word comprehension we could systematically examine more potential moderators. Each study included in this meta-analysis assessed word learning for a specified number of individual target words using a comprehension task similar to the Peabody Picture Vocabulary Task (Dunn & Dunn, 2015). We considered how who read the story, reading style, child's age, tokens, story repetitions and word type moderate word comprehension. Our aim was to provide helpful guidance for best practices, both for experimental design and for shared storybook reading in naturalistic settings.

Method

Literature search and inclusion criteria

We conducted a systematic search of the online databases British Education Index, Education Resources Information Center, PsycInfo, Scopus and Web of Knowledge. Figure 1 shows a schematic of the search strategy. We used the following search terms, alone and in combination: *Child*, Language Acquisition, Picture Book, Picturebook, Story, Story Book, Storybook, Vocabulary,* and *Word Learning*. In addition, we used relevant mailing lists and personalized emails to contact key researchers in the field, asking for details of relevant studies to include (published and unpublished). Finally, reference lists from relevant reviews were searched to identify additional studies. To be included in the analysis, studies were required to meet several criteria (see Figure 1 for our selection strategy), consistent with PICOS guidelines (Liberati et al., 2009).

We included studies published after Elley (1989): a seminal shared storybook reading study. Thus, we included studies dated from 1990 until the original search date of January 2015 and a follow-up search date of May 2017. The search procedures were identical. Included study designs provided a measure of children's word learning following an adult reading one or more storybooks. As such, most of the studies tested children who were not yet in formal schooling though some studies did involve teachers reading to school-aged children. All participants were typically-developing children and not specifically multi-lingual. Studies where fewer than half of the children were second language learners were retained if investigating second language acquisition was not the focus of the research. Studies on communication disorders were retained if there was at least one typically-developing control group, and in these cases only control groups were included in the meta-analysis. Storybook readings could have been single or multiple events, and adult readings included live or pre-recorded stories from real books or displayed on computer screens, provided the images were static.

Comprehension tests using 4-alternative forced-choice trials like those used on the Peabody Picture Vocabulary Test (e.g., PPVT-4, Dunn & Dunn, 2007) are widely employed to test children's comprehension of target words and provide a comparable standard across studies. This measure involves asking children to select the picture that best reflects a given word from an array of four pictures. Where children do not know the correct answer, the likelihood of being correct by chance is 25%. We analysed only studies that use this testing method. To ensure effect size comparisons were meaningful, we excluded studies that used arrays of more than four pictures (e.g., Ewers & Brownson, 1999), or required children to provide definitions (e.g., Biemiller & Boote, 2006; Leung, 2008) because chance levels would differ.

Experimental designs varied across the studies we included, but followed similar patterns, generally related to the kinds of target words that they included. There were three kinds of target words: words that were confirmed unknown using a pre-test or pilot to confirm their novelty (e.g., *forlorn*, Beck & McKeown, 2007), sophisticated words selected to ensure their relative novelty for the target age range (e.g., *departed* for 4- or 6-year-olds, Houston-Price, Howe, & Lintern, 2014), or completely novel pseudo-words (e.g., *manu*, Horst et al., 2011). Studies that used confirmed unknown words (e.g., *forlorn*) included pre-tests or pilots to determine baseline word knowledge before storybook reading (e.g., Abel & Schuele, 2013; Wilkinson & Houston-Price, 2013) and therefore did not include no storybook reading control groups. Similarly, studies that used completely novel pseudo-words (Williams & Horst, 2014) used novel words so that any word learning could be directly attributed to the storybook intervention. Using completely novel words that children would have no knowledge of before the study effectively means a pre-test score of 0 and turns the test scores into post-test scores because only storybook reading can account for any knowledge gained. Thus, studies that used completely novel target words also did not include no storybook reading control groups. Other studies included control groups but the task for the control groups varied. For example, children in the control groups were exposed to different books or different reading techniques (e.g., Horst et al., 2011; Mandel, Osana & Venkatesh, 2013). In these cases we opted to treat so-called control groups as separate effects sizes because the children still received storybook exposures and the comprehension tests met our design criteria for inclusion. This keeps the calculation method for the individual effect sizes consistent across research designs.

Finally, sufficient information to compute the relevant effect sizes was required. To this end, we retained studies for which a copy of the paper (or in the case of unpublished studies, the original dataset) was obtained. We made an exception for pre/post-test correlations as these were provided in so few papers. We emailed 12 authors (15 journal articles) where information was missing. Three authors replied, but only one was able to provide the correlations we required. There were an additional three authors for whom no contact details could be found. Note, as a meta-analytic review The University of Sussex did not require additional ethical approval for this research.

Analysis strategy

Studies in the meta-analysis used either pre-post designs, or designs in which the target words were novel (and so no pre-test was necessary). Although it would be common to quantify effects in terms of a standardized change score (such as Cohen's *d*), such measures were problematic because (1) when novel words were used the mean and standard deviation of pre-test scores would be 0; and (2) in many studies there was no information about the correlation between pre- and post-test scores, therefore *d* would need to be computed making assumptions about the size of this relationship.

Therefore, the meta-analysis was conducted on the change in raw test scores. That is, for each study the effect size was the number of new words learned between the pre- and post-test (i.e., the difference between pre- and post-test scores). Where no pre-test was conducted a pre-test score of 0 was assumed, note this includes studies that used completely novel words (e.g., *manu*) as described above. The meta-analysis was conducted using the metafor software package (Viechtbauer, 2010) for R (R Core Team, 2015). Given that meta-analyses of raw score differences is relatively uncommon, we conducted an alternative analysis that expressed the pre-to-post change in test scores as Cohen's *d* (see above for caveats about this approach). The data and results of this alternative analysis are available at https://osf.io/rxbdz/.

Coding. Each study was assigned a unique study code, and effect sizes within studies were also numbered. A coding frame was agreed and primary coding was completed by the first author. We coded general study descriptors such as publication type and dates and participant age and test type (e.g., 4-alternative forced-choice).

For potential moderators we coded year of publication, reader (e.g., experimenter, teacher, parent or combination), book genre (non-fiction or fiction), book type (computer or paper) and illustration type (photos or cartoons). For moderators related to testing, we coded the delay between storybook reading and testing as a continuous variable. We also included the number of story repetitions, tokens and the proportion of nouns used as target vocabulary. We chose to calculate this as a proportion of overall target words because this enabled us to retain the maximum number of studies for this moderator, given that some studies combined nouns and verbs (e.g., Gallingane, 2010; Zipoli, Coyne, & McCoach, 2010).

Finally, we also calculated the total number of "words tested," which is not necessarily the number of intended target words. For studies using completely novel words (e.g., *manu*), all target words counted as words tested because children could not have any knowledge of these words at the start of the study. However, for studies using confirmed unknown words (e.g., *forlorn*) pre-test scores sometimes suggested *a priori* knowledge of some of the target words. In these cases, we subtracted the pre-test scores from the total number of target words to provide a more conservative number of words tested. For example, Abel and Schuele (2014) intended their study to include 15 target words. At pre-test, children correctly identified on average 6.51 words (*SD* = 2.80), suggesting they already knew some of the target words. In this case, we subtracted the 6.51 words already known from the 15 intended target words. This resulted in 8.49 "words tested." Put another way, the unknown "words tested" was the difference between the intended number of target words and those children had correctly identified on the pre-test before the storybook intervention.

To assess reliability of moderator and effect size calculations and coding studies to be excluded, a second coder randomly selected 15% of both excluded and included studies. Agreement of 100% was reached on which studies were to be excluded. For exclusion reasons and for moderator and effect size data, Cohen's Kappas ranged from .95 to 1. Final codes used for analysis were reached by agreement.

Analysis. The final data set consisted of 38 unique studies, contributing 110 effect sizes. Some studies included multiple effect sizes. For example, because they tested multiple age-groups (e.g. Wilkinson & Houston-Price, 2013) or different reading conditions (e.g., McLeod & McDade, 2011). Studies which contribute multiple effect sizes not only violate the assumption of independence of effect sizes but can distort population effect sizes, because a single study can contribute several times to the overall effect size calculation (Field, 2015). For this reason, we used the *rma.mv()* function with restricted maximum-likelihood estimation in metafor to perform a multi-level meta-analysis, in which effect sizes (level 1) are nested within studies (level 2). We allowed effect sizes to vary across studies (random effects) and entered moderators as fixed effects. We calculated robust 95% BCa confidence intervals around the parameter for each moderator using 1000 bootstrap samples because the number of effect sizes, *k*, was generally small. We also repeat the analysis with 3 outliers removed (see Table 1).

Moderators.

Reader, reading style and child age. "Reader" (experimenter, teacher) varied across studies. In one case, the experimenter was also the teacher, and in one case parents read, so these two cases were excluded from the reader moderator analysis. We measured the moderator "dialogic reading" categorically. Specifically, we classified reading styles as dialogic if the reader added something to a verbatim text reading. For example, reading styles with pointing, repetitions via extra-textual speech, or the addition of dictionary definitions as they read for target vocabulary were classified as dialogic. Dialogic and non-dialogic styles accounted for 49 and 61 of the effect sizes, respectively. Child's "age" was measured in months, based on best estimates from data provided and varied from 35 to 122 months (M = 62.71, median = 55.94, SD = 23.13).

Experimental design. The number of "story repetitions" were provided in 104 cases and ranged from a single reading of a story through to seven readings (M = 2.19, SD = 1.04). "Words tested" measured the number of words children did not know and could therefore learn. The number of words tested ranged from 2 to 34.84 words (M = 7.48, SD = 5.11). "Tokens" refers to the number of word exposures. Not all studies control for tokens. For example, in some dialogic studies dictionary definitions (e.g., Penno, Wilkinson, & Moore, 2002), or repetitions (Leung, 2008; Reese & Cox, 1999) are used freely and therefore tokens are not reported by the original authors. Tokens were provided for 83 effect sizes and ranged from 1 to 12 tokens (M = 6.01, SD = 4.37). The "story to test interval" was measured in hours, based on the best estimate using the data provided. For example, where authors reported a delay of 3 days, we calculated this as 72 hours. Story to test interval varied from immediate to 10 weeks (M = 363.75 hours, or 2 weeks, 1 day, SD = 448.78 hours, or 2 weeks, 4 days).

Word type. Several studies report different rates of word learning for nouns and verbs (e.g., McLeod & McDade, 2011), thus we aimed to include this as a moderator. Many studies included both nouns and verbs, and some do not report separate scores for nouns and verbs, therefore, we analyzed this as a continuous variable, "proportion of nouns." For each effect size we calculated the proportion of target words that were nouns (M = .91, SD = 0.20). For 44 effect sizes we could not calculate proportion of nouns because the authors did not specify how many of the target words were nouns. The categorical moderator "word novelty" reflects the use of novel words (e.g., *manu*, $k = 38$) or real words, which are either sophisticated known words (e.g., *departed,* $k = 16$), or confirmed unknown (e.g., *forlorn,* $k = 56$).

Results

Word Comprehension from Storybooks

A total of 110 effect sizes from 38 studies with 2,455 children contributed to these analyses. Figure 2 presents the forest plot (Lewis & Clarke, 2001) of effect sizes for each study. Note, that the models fitted included multiple effect sizes within some studies whereas for the forest plot the average effect size within a study is displayed. Comprehension test studies report a positive effect of shared storybook reading on word learning, $k = 110$, raw change = 3.025 words [2.622, 3.366], $p < .001$. Overall, children learned 45% (SD = 24%) of the words to which they were exposed. Studies in this sample were also significantly heterogeneous, $Q(109) = 10019.576$, $p < .001$, suggesting variability in the studies which may be accounted for by moderator variables (summarized in Table 1).

Moderator Analyses

Studies ranged in the number of target words and the number of words tested. Critically, this could mean that a raw score of 4 words correct could be as high as 75% accuracy in a study that included 6 target words, but could be as low as 25% for a study that included 12 target words. To adjust moderators for this we included the total number of target words indicated by each study's author(s) in each of the models. We chose the total number of target words rather than number of *words tested* because all studies provided the number of intended target words whereas words tested could only be calculated for studies that included pre-test measures or completely novel words.

Words tested was a significant predictor of word comprehension, although the effect size was very small. Despite another relatively small effect size for tokens as a moderator, 95% BCa confidence intervals suggested the real effect size could be zero. For comparison we provide details of both the moderators “target words” and “words tested,” modelled in isolation at the top of Table 1. All subsequent moderation analyses include the “target words” variable in addition to the moderator of interest.

Reader and Reading Style. Whether a teacher or experimenter read the book was not a significant moderator of word comprehension (see Table 1). The use of dialogic reading styles, however, was a significant moderator. Specifically, using dialogic reading styles (k = 49) results in children learning 1.22 more words than non-dialogic reading (k = 61) after adjusting for number of target words tested. Thus, when it comes to learning words from storybooks, *how* stories are read is more important than *who* reads them.

Despite a p-value of less than .05, the 95% BCa bootstrap confidence interval for age when adjusted for number of target words included zero. Assuming that this CI is one of the 95% that contains the population value, this means that the population effect of age on word leaning could be zero. In any case, the estimate of the effect revealed a very small effect (Table 1): 0.06 of a word, thus, regardless of the significance, the effect of age on the ability to learn words from shared storybook reading appears trivial.

Experimental Design. Despite finding a moderate effect size for tokens, BCa confidence intervals included 0 suggesting the true population effect could be 0. The very small estimate for the number of story repetitions adjusted for the number of target words suggests story repetitions are a predictor of word learning, although widely spaced BCa confidence intervals which included 0 suggest the true population effect could be 0. Because precise numbers of tokens were not always reported (particularly in studies using dialogic techniques), we also repeated this analysis with only those studies that provided a precise number of tokens. For the 61 cases with story repetitions ranging from 1 to 7 (M = 2.28, SD = 1.11) we found a similar estimate of words learned (0.151, [-0.138, 0.402], p < .001) when adjusting for the number of target words. Once again though, BCa confidence intervals included 0. In addition, the estimate for story to test interval was very small and confidence intervals suggest the population effect is likely to be 0. Overall this suggests a minimal effect of story to test interval on word comprehension.

Word Type. For studies providing a separate breakdown of performance for nouns and verbs the mean number of nouns learned was 2.77 (SD = 1.73, range = 0.76 – 8.28, k = 40) out of a mean of 6.29 nouns presented and the mean number of verbs learned was 3.10 (SD = 0.64, range = 2.64 - 3.55, k = 2) out of a mean of 5.00 verbs presented. Remaining effect sizes included combinations of nouns, verbs or adjectives. Despite a large effect size, the proportion of nouns among the target words did not significantly predict word learning. Confidence intervals were wide and contained 0, suggesting potentially little difference in word learning from stories that included only nouns or a mix of word types.

Word novelty, i.e., the kind of target word used (e.g., *forlorn*, *departed*, *manu*), predicted word comprehension when adjusted for the total number of target words. Specifically, using confirmed unknown words (e.g., *forlorn*, k = 56) results in significantly fewer words learned than tests, which use novel words (e.g., *manu*, k = 38). Although estimates using sophisticated words (e.g., *departed*, k = 16) suggested significantly greater word learning than using novel words, wide confidence intervals including 0 suggest this effect could be 0.

Outliers

We assessed the sample for outliers, using residual Q-Q plots and investigated the influence these studies had over the wider sample, using leverage plots (e.g., Cook's distance, hat values etc.). From these we identified three possible outliers in our sample (plots are available for inspection at https://osf.io/rxbdz/). These studies had standardized residuals over or close to 3, and had noticeably higher values than other studies on their standardized residuals, Cook's distances and dffit statistics. These three studies (five effect sizes) included design features that our analyses demonstrate should lead to high rates of word learning from storybooks. Specifically, each of these studies included completely novel words and two studies (four effect sizes) used dialogic reading techniques. A closer inspection of the procedure for one study in particular revealed that children received instructions that they would hear some "silly words… and they should help the examiner figure out what the words meant" (Horohov & Oetting, 2004, pp. 52). This may have focused children's attention on the words, by making word learning the explicit objective of the task, which is not typical in the procedures of other studies.

We repeated our analysis with these three studies removed and our results are depicted in the right-hand columns of Table 1 for comparison. The removal of these studies did not substantively change the findings for the main analysis, although it reduced the estimate to a more moderate 2.571 words learned, [2.237, 2.856], $p < .001$. Individual moderator analyses were largely unaffected by removal of these cases, except in the case of tokens, the proportion of nouns, and word novelty. The BCa confidence intervals for tokens, after adjusting for the number of target words used, changed to become more precise and no longer included 0 after removal of outliers. This suggests more tokens, i.e., more exposures to target words, leads to better word learning. The proportion of nouns, after adjusting for the number of target words, still produced a large effect size estimate, but confidence intervals no longer included 0, suggesting that although the estimate still lacked precision, the true population effect is not likely to be 0. After adjusting for the number of target words, word novelty, in particular, the use of sophisticated, but not pre-tested words, were associated with higher word learning performance than when compared with novel words. Confidence intervals suggest this could be a population wide effect.

Publication Bias

Figure 3 shows the funnel plots for the studies in this meta-analysis. There is a clear sparsity in the lower right and no effects in the lower left (i.e. truncation below zero) indicative of one-tailed, selection bias. In particular, the plot shows a lack of studies in the lower left quadrant. This pattern is entirely consistent with the conclusion that studies with small samples (and, therefore, larger standard errors) that produced small effects are missing from the published literature. Such studies would be unlikely to contain significant effects (both because the effects are small, and the studies are underpowered to detect them). This pattern is consistent with publication bias: small changes in small samples will not be significant and will not be published, small changes in large samples will be significant and will be published. Larger effects (effects towards the right of the plot) are published regardless of sample size (as you'd expect because larger effects will be significant irrespective of sample size).

Qnorm plots of residuals (available at https://osf.io/rxbdz/) suggest publication bias, but for a population with a non-zero effect size (Wang & Bushman, 1998). To quantify the likely effect of publication bias, a sensitivity analysis based on Vevea and Woods (2005) was conducted. We modelled a scenario in which nonsignificant studies ($0.05 < p < 1$) were about half as likely to be published as significant studies (i.e. a fairly severe situation). The overall population effect size estimate changed from 3.027 to 2.79[1]. Even correcting for severe selection bias the overall population effect size estimate was reduced only modestly. This sensitivity analysis suggests that our estimates would not be radically reduced by the inclusion of unpublished studies based on our assumptions about the likely pattern of selection bias.

We also conducted additional analyses to investigate the relationship between study variance and effect size. Mediation analyses suggested that story to test interval, ($p = .022$), target words ($p = .078$), and the proportion of nouns ($p = .059$) provide some mediation. Further information about these analyses is provided at https://osf.io/rxbdz/

Discussion

In addition to being a highly enjoyable activity for both children and adults, shared storybook reading supports reading skills (Bus et al., 1995; Niklas, Cohrssen, & Tayler, 2016; Sénéchal & LeFevre, 2002) and later vocabulary development (Elley, 1989; Robbins & Ehri, 1994; Sénéchal & Cornell, 1993) by providing a richer source of linguistic input than conversation alone (Mol et al., 2008; Montag et al., 2015). This is important for child development because a child's vocabulary size predicts later academic performance (Snow, Griffin, & Burns, 2005), behavioural regulation (Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015) and even criminal convictions (Murray, Irving, Farrington, Colman, & Bloxsom, 2010). We used multi-level meta-analysis to estimate the population effects of shared storybook reading on word learning and to investigate the factors moderating this relationship from studies, which together included 2,455 participants. Overall, we found children were able to comprehend just under half of the new words to which they were exposed which is consistent with other reports in the literature (see, e.g., Biemiller & Boote, 2006).

Overall, learning is the product of embodied children interacting with other people and their environments (Thelen & Smith, 1994). In highly naturalistic situations, such as shared storybook reading, there will be multiple factors simultaneously influencing children's learning, including what the child is currently seeing and hearing (e.g., the attributes of a given storybook) and the child's previous knowledge (i.e., the child's own developmental history interacting with books). Although we were keen to understand these important interactions, we limited ourselves to simple moderator analyses in the current meta-analysis. There were too many explicit variables across studies to perform multiple moderator analyses for every possible combination of factors that could interact with each other while maintaining statistical power. The original papers discuss their own interactions. In addition, we noticed orthogonal patterns in the literature. For example, studies using completely novel words (e.g., *manu*) tended to also include very few target words. Likewise, studies using dialogic reading styles tended to not control for tokens. Below we review our findings and discuss the orthogonal patterns we encountered as well as highlight areas that need more research.

Word Learning From Storybooks: What Really Matters

How You Read. Our results indicate reading style and use of dialogic techniques (such as pointing, providing definitions or asking children questions as you read) significantly influences the number of new words children learn from shared storybook reading. In fact, our results suggest that, after adjusting for the number of target words, the use of dialogic styles increases word learning by more than one word per child. Many different dialogic techniques were employed across the studies reviewed. Techniques included describing pictures (Reese & Cox, 1999), providing dictionary definitions before readings (Coyne et al., 2004), asking questions during readings (Blewitt, Rump, Shealy, & Cook, 2009; Walsh & Blewitt, 2006) or incorporating music (Joyce, 2012). Although we suspect there may be different effects across techniques (e.g., Ard & Beverly, 2004; Reese & Cox, 1999) we did not investigate the effects of these different styles separately as we had already planned to investigate so many individual moderators.

Tokens. The moderating effects of tokens on word comprehension when adjusted for the number of target words, were small but precise, indicating the importance of multiple exposures for word learning. Increasing the number of tokens provides children with greater opportunity to learn and consolidate new words (Samuelson, 2002; Woodward, Markman, & Fitzsimmons, 1994). This will surprise few people—after all, word learning experiments have consistently manipulated tokens to improve word learning both directly (e.g., Blewitt et al., 2009) and indirectly (e.g., through storybook repetitions; Sénéchal, 1997; Sénéchal, Thomas, & Monker, 1995). Similarly many dialogic reading techniques are specifically aimed at increasing tokens, for example, by asking children to repeat words. Our analysis (after removal of outliers) provides evidence of a true population effect. This is encouraging because increasing tokens is a remarkably simple intervention, which can be used by parents and teachers alike with minimal specific training.

The Number of Words Tested. Overall, we found some evidence of moderating effects from the number of new words tested. The number of words tested was calculated by subtracting the pre-test scores (i.e., words children already knew) from the number of intended target words. These effects remained remarkably stable, even after removal of outliers. In some studies only a subset of the words deemed difficult enough to be selected as target words on the test trials were actually used as target words (e.g., Vuattoux, Japel, Dion, & Dupéré, 2014; Weisberg et al., 2015). However, in these cases children still encountered all of the potentially new words during storybook reading (e.g., a storybook may contain the not-fully known word *departed,* but that word may not be on the researcher's list of target words). Thus, children may be working to learn additional new words beyond those that the researchers intended to test. That is, the number of target words is not necessarily an indication of the level of difficulty of a story. If the entire story contains a high number of unknown words—even if some are not explicitly tested—then children's overall word learning could be underestimated. Therefore, researchers should take care to ensure non-target words are still well-matched to children's abilities.

Word Learning From Storybooks: What Matters Less

Who Reads. Children's word comprehension does not appear to be influenced by whether a story is read by a familiar teacher or a researcher the child has just met. For experimental design, the lack of an effect for word comprehension is encouraging as it suggests results obtained in the lab may generalize to naturalistic settings.

Child's Age. Hundreds of studies within educational and developmental psychology, and within the language acquisition literature more generally, demonstrate that older children have more proficient word learning skills than younger children (e.g., Bion et al., 2013). Therefore it should follow that word learning increases with age, however, this was not what we found in the shared storybook reading literature. In our analysis any moderating effects of age were trivial. One explanation for why age does not moderate word learning from storybooks is because experiments are designed with a specific participant age-range in mind and researchers already take into account an appropriate level of challenge. Indeed, a closer examination of our database indicates that storybooks read to older children include more target words ($r = .24$, $p = .01$) and fewer tokens ($r = -.38$, $p < .001$) than those read to younger children. This indicates that researchers *are* taking overall book difficulty into account when they select storybooks for use in research. In addition, other factors that correlate with age but not available to us, such as pre-experimental vocabulary size, or amount of experience with storybooks, may also contribute to word learning success.

Story Presentation to Test Interval. For word comprehension, we found no significant moderating effects of story to test interval. This finding is unexpected in light of other studies demonstrating that differences in stimuli presentation effects both short- and long-term word learning success (e.g., Vlach, Ankowski, & Sandhofer, 2012). In our sample, the delay between the reading and testing varied from immediate testing to testing after 10 weeks, but with some noticeable gaps. For example, there were no studies testing delays of between 1 and 2 weeks. Thus, our finding may be affected by missing data. In addition, some studies in our sample assessed word learning only once, some twice and some three times, but no studies assessed word learning more frequently than three times. Adolph, Robinson, Young, and Gill-Alvarez (2008) argue that more frequent sampling is beneficial for discovering the shape of developmental change. At first glance, testing more frequently sounds like an ideal recommendation for research in word learning from shared storybooks.

However, every time children are tested they are re-exposed to the target words and/or definitions. That is, they are given another opportunity to encode (Munro, Baker, McGregor, Docking, & Arciuli, 2012) and learn (Horst & Simmering, 2015), which then makes it difficult to determine how much of their word learning is due to their initial storybook exposure.

Word Learning From Storybooks: Topics for Future Research

Story Repetitions. Repeated readings of the same storybooks benefit word learning, even after controlling for tokens (Biemiller & Boote, 2006; Horst et al., 2011; McLeod & McDade, 2011). We found a small effect of repetition on word comprehension, but confidence intervals suggest a lot of variability in these studies, and the true effect could be 0. Story repetition may reflect another orthogonal pattern in word learning from shared storybook reading, with studies that include repeated readings testing fewer words. Alternatively, repetition effects could be stronger at some ages than others. For example, Horst et al. (2011); McLeod and McDade (2011); and Sénéchal (1997); Damhuis, Segers, and Verhoeven (2015) all found robust effects of story repetition with 3- to 5-year-old children. These explanations are not mutually exclusive, but they do suggest further investigation into the effects of storybook repetition on word learning is needed.

Word Novelty. Target words were divided into three categories: real words that were either confirmed unknown (e.g., *forlorn*), deemed too sophisticated for the age-range of children being tested (e.g., *departed*) or completely novel, plausible pseudo-words (e.g., *manu*). Word novelty was a strong moderator of word comprehension. Most interestingly, word learning effects differed greatly for each word type. For example, even after adjusting for the number of target words, children correctly identified more than one whole extra word in tests using novel words (e.g., *manu*) than confirmed unknown (e.g., *forlorn*). However, children correctly identified more than one whole word more than that in tests using sophisticated words (e.g., *departed*) than novel words.

These results are surprising, but we believe this highlights the impact of methodological differences across studies. One possible explanation for why studies using novel words resulted in fewer words learned than those using sophisticated words is that of pre-experimental control. All words are novel at some point, but repeated encounters to new words results in a gradual transition from unknown to known. Word learning requires multiple exposures to a word (e.g., McMurray et al., 2012). Children tested with real words may have some pre-experimental exposure to these words, which is not measured, therefore providing an advantage when tested over entirely novel vocabulary. Novel words control for children's pre-experimental exposure to target words because children definitely have not encountered these elsewhere. However, this explanation would also predict that studies using novel words would result in fewer words learned than those using confirmed unknown words, but this was not the case. Though, as noted above, studies using novel words typically included fewer target words overall. This leads us to believe this is an important consideration in research design, particularly where comparisons are made between studies using words of varying degrees of novelty.

Word Type. Overall, we found 2- to 10-year-old children could learn approximately 2-5 new words from shared storybook reading (approximately 2.77 nouns, 3.10 verbs). The proportion of nouns tested suggested this was a strong moderator of word learning, however, widely spaced confidence intervals included 0, suggesting the true effect could be zero. Interestingly, with outliers removed, nouns were still not learned better than verbs. There is abundant evidence that nouns are easier to learn than verbs (for a review see Golinkoff & Hirsh-Pasek, 2008), but our analyses showed conflicting evidence of this in the shared storybook reading literature. Differences in word learning effects for nouns and other word types is a worth-while topic, which merits further investigation.

Conclusions

Our review integrates studies from education, developmental psychology and the control groups in communication disorders research to quantify the effect of shared storybook reading on children's word comprehension. Although large positive word learning effects were expected, the unique value here was in gaining an understanding of the many factors moderating these effects.

This meta-analysis highlights several important variables, which truly affect the quality and quantity of word learning from shared storybook reading. Reading style is of paramount importance: dialogic reading styles that encourage additional interaction with the text significantly improve word learning. Both the number of new words introduced and how often they are heard are clearly important. The effects of story repetitions, word novelty and word type are more variable. Future research should explore these factors. Overall, the insights gained from this meta-analysis provide valuable guidance to researchers investigating word learning from shared storybook reading, and to parents, teachers or speech and language therapists wishing to provide the best possible learning environment for children.

Footnotes

[1] To do this analysis it was necessary to aggregate effect sizes within studies so that each study produced only 1 effect size. This differs from the multi-level approach used in the main analysis, which explains the difference in the unadjusted value here and that reported earlier.

References

References marked with an * indicate studies included in the meta-analysis.

*Abel, A. D., & Schuele, C. M. (2014). The influence of two cognitive-linguistic variables on incidental word learning in 5-year-olds. *Journal of Psycholinguistic Research, 43*(4), 447-463. doi: 10.1007/s10936-013-9264-4

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change. *Psychological Review, 115*(3), 527-543. doi: 10.1037/0033-295x.115.3.527

Aram, D., & Besser, S. (2009). Early literacy interventions: Which activities to include? At what age to start? And who will implement them? *Infancia y Aprendizaje / Journal for the Study of Education and Development, 32*(2), 171-187. doi: 10.1174/021037009788001806

*Ard, L. M., & Beverly, B. L. (2004). Preschool word learning during joint book reading: Effect of adult questions and comments. *Communication Disorders Quarterly, 26*(1), 17-28. doi: 10.1177/15257401040260010101

Barratt-Pugh, C., & Rohl, M. (2015). Better beginnings has made me make reading part of our everyday routine: Mothers' perceptions of a family literacy program over four years. *Australasian Journal of Early Childhood, 40*(4).

*Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251-271. doi: 10.1086/511706

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44-62. doi: 10.1037/0022-0663.98.1.44

Bion, R., A. H.,, Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18,24, and 30 months. *Cognition, 126*, 39-53. doi: 10.1016/j.cognition.2012.08.008

Blake, J., Macdonald, S., Bayrami, L., Agosta, V., & Milian, A. (2006). Book reading styles in dual-parent and single-mother families. *British Journal of Educational Psychology, 76*(Pt 3), 501-515. doi: 10.1348/000709905X49719

*Blewitt, P. & Langan, R., (2016). Learning words during shared book reading: The role of extratextual talk designed to increase child engagement. *Journal of Experimental Child Psychology, 150*, 404-410. doi: 10.1016/j.jecp.2016.06.009

*Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology, 101*(2), 294-304. doi: 10.1037/a0013844

*Brett, A., Rothlein, L., & Hurley, M. (1996). Vocabulary Acquisition from Listening to Stories and Explanations of Target Words. *The Elementary School Journal, 96*(4), 415-422.

Bus, A. G. (2001). Joint caregiver-child storybook reading: A route to literacy development. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 179-191). New York, NY: The Guildford Press.

Bus, A. G., van Ijzendoorn, M. H., & Pellegrini, A. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65*, 1-21. doi: 10.3102/00346543065001001

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264-293). Cambridge, MA: MIT Press.

*Chen, Y., & Liu, H.-M. (2014). Novel-word learning deficits in Mandarin-speaking preschool children with specific language impairments. *Research in Developmental Disabilities, 35*(1), 10-20. doi: 10.1016/j.ridd.2013.10.010

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*(3), 145-162. doi: 10.1207/s15327035ex1203_3

*Damhuis, C. M. P., Segers, E., & Verhoeven, L. (2015). Stimulating breadth and depth of vocabulary via repeated storybook readings or tests. *School Effectiveness and School Improvement, 26*(3), 382-396. doi: 10.1080/09243453.2014.965181

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Peabody picture vocabulary test* (4th ed.). Minneapolis, MN: Pearsons Assessments.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*(2), 174-187. doi: 10.2307/747863

*Evans, M. A., & Saint-Aubin, J. (2013). Vocabulary acquisition without adult explanations in repeated shared book reading: An eye movement study. *Journal of Educational Psychology, 105*(3), 596-608. doi: 10.1037/a0032465

Ewers, C. A., & Brownson, S. M. (1999). Kindergarteners' vocabulary acquisition as a function of active vs. passive storybook reading, prior vocabulary, and working memory. *Reading Psychology, 20*(1), 11-20. doi: 10.1080/027027199278484

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development *Monographs of the society of research in child development.* (Vol. 59).

*Flack, Z. M, Gambi, C. & Horst, J. S. (2016). *[Raw data and study description]*. Unpublished raw data.

*Flack, Z. M. & Horst, J. S. (2016). *[Raw data and study description]*. Unpublished raw data.

*Flack, Z. M. & Horst, J. S. (2017). Two sides to every story: children learn words better from one storybook page at a time. *Infant and Child Development.* Advance online publication. doi: 10.1002/icd.2047

Field, A. P. (2015). Dread returns to mega-silly one. *Health Psychology Review, 9*(1), 15-20. doi: 10.1080/17437199.2013.879198

Gallingane, M. C. (2010). *Effects of read-aloud strategies on young children's vocabulary learning.* Retrieved from ProQuest Dissertations Database.

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences, 12*(10), 397-403. doi: 10.1016/j.tics.2008.07.003

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*(1), 75-90.

Hindman, A. H., Connor, C. M., Jewkes, A. M., & Morrison, F. J. (2008). Untangling the effects of shared book reading: Multiple factors and their associations with preschool literacy outcomes. *Early Childhood Research Quarterly, 23*(3), 330-350. doi: 10.1016/j.ecresq.2008.01.005

*Horohov, J. E., & Oetting, J. B. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics, 25*(01), 43-65. doi: 10.1017/S0142716404001031

*Horst, J. S. (2015a). *[Raw data and study description]*. Unpublished raw data.

*Horst, J. S. (2015b). *[Raw data and study description]*. Unpublished raw data.

*Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: contextual repetition promotes word learning from storybooks. *Frontiers in Psychology, 2*. doi: 10.3389/Fpsyg.2011.00017

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*(2), 128-157. doi: 10.1080/15250000701795598

Horst, J. S., & Simmering, V. R. (2015). Category learning in a dynamic world. *Frontiers in Psychology, 6*, 46. doi: 10.3389/fpsyg.2015.00046

*Houston-Price, C., Howe, J. A., & Lintern, N. J. (2014). Once upon a time, there was a fabulous funambulist...: What children learn about the "high-level" vocabulary they encounter while listening to stories. *Frontiers in Psychology, 5*, 75. doi: 10.3389/fpsyg.2014.00075

Huttenlocher, J. (1974). The origins of language comprehension. In R. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium*. Oxford, England: Erlbaum.

Joyce, M. F. (2012). *Vocabulary acquisition with kindergarten children using song picture books.* (Doctoral Dissertation), Northeastern University, Boston, Massachusetts.

*Justice, L. M. (2002). Word exposure conditions and preschoolers' novel word learning during shared storybook reading. *Reading Psychology, 23*(2), 87-106. doi: 10.1080/027027102760351016

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*(1), 17-32. doi: 10.1044/0161-1461(2005/003

*Kaefer, T., Neuman, S. B., & Pinkham, A. M. (2014). Pre-Existing Background Knowledge Influences Socioeconomic Differences in Preschoolers' Word Learning and Comprehension. *Reading Psychology*, 1-29. doi: 10.1080/02702711.2013.843064

Lawson, K. (2012). The real power of parental reading aloud: Exploring the affective and attentional dimensions. *Australian Journal of Education, 56*(3), 257-272.

Leung, C. B. (2008). Preschoolers' acquisition of scientific vocabulary through repeated read-aloud events, retellings, and hands-on science activities. *Reading Psychology, 29*(2), 165-193. doi: 10.1080/02702710801964090

Lewis, S., & Clarke, M. (2001). Forestplots: Trying to see the wood and the trees. *British Medical Journal, 322*(7300), 1479-1480. doi: 10.1080/02702710801964090

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, J. P., Ioannidis, A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med, 6*(7). doi: 10.1371/journal.pmed.1000100

*Mandel, E., Osana, H. P., & Venkatesh, V. (2013). Addressing the Effects of Reciprocal Teaching on the Receptive and Expressive Vocabulary of 1st-Grade Students. *Journal of Research in Childhood Education, 27*(4), 407-426. doi: 10.1080/02568543.2013.824526

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300-335. doi: 10.3102/0034654310377087

*McLeod, A. N., & McDade, H. L. (2011). Preschoolers' incidental learning of novel words during storybook reading. *Communication Disorders Quarterly, 32*(4), 256-266. doi: 10.1177/1525740109354777

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning as the interaction of onine referent selection and slow associative learning. *Psychological Review, 119*(4), 831-877. doi: 10.1037/a0029872

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development, 19*(1), 7-26. doi: 10.1080/10409280701838603

Montag, J., L., Jones, M. N., & Smith, L., B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science, 26*(9), 1489-1496. doi: 10.1177/0956797615594361

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development, 86*(5), 1351-1370. doi: 10.1111/cdev.12398

Munro, N., Baker, E., McGregor, K. K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology, 3*(41). doi: 10.3389/fpsyg.2012.00041

Murray, J., Irving, B., Farrington, D. P., Colman, I., & Bloxsom, C. A. (2010). Very early predictors of conduct problems and crime: results from a national cohort study. *Journal of Child Psychology and Psychiatry, 51*(11), 1198-1207. doi: 10.1111/j.1469-7610.2010.02287.x

Niklas, F., Cohrssen, C., & Tayler, C. (2016). The Sooner, the Better. *SAGE Open, 6*(4), 215824401667271. doi: 10.1177/2158244016672715

Ninio, A. (1980). Picture-book reading in mother-infant dyads belonging to two subgroups in Israel. *Child Development*, 587-590.

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*(1), 23-33. doi: 10.1037/0022-0663.94.1.23

Pillinger, C., & Wood, C. (2014). Pilot study evaluating the impact of dialogic reading and shared reading at transition to primary school: early literacy skills and parental attitudes. *Literacy, 48*(3), 155-163.

*Pinkham, A. M., Kaefer, T., & Neuman, S. B. (2014). Taxonomies Support Preschoolers' Knowledge Acquisition from Storybooks. *Child Development Research, 2014*, 1-10. doi: 10.1155/2014/386762

*Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice, 25*(3), 110-123. doi: 10.1111/j.1540-5826.2010.00309.x

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

*Ralli, A. M., & Dockrell, J. E. (2005). Multiple measures of assessing vocabulary acquisition: Implications for understanding lexical development. *Psychology: The Journal of the Hellenic Psychological Society, 12*(4), 587-603.

*Ramachandra, V., Hewitt, L. E., & Brackenbury, T. (2011). The relationship between phonological memory, phonological sensitivity, and incidental word learning. *J Psycholinguist Res, 40*(2), 93-109. doi: 10.1007/s10936-010-9157-8

Reese, E., & Cox, A. (1999). Quality of adult book reading affects children's emergent literacy. *Developmental Psychology, 35*(1), 20-28.

*Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86*(1), 54-64.

Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15–20-month-olds. *Developmental Psychology, 38*(6), 1016–1037. doi: 10.1037//0012-1649.38.6.1016

Schwartz, J. I. (2004). An observational study of mother/child and father/child interactions in story reading. *Journal of Research in Childhood Education, 19*(2), 105-114. doi: 10.1080/02568540409595058

*Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language, 24*(1), 123-138.

*Sénéchal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28*(4), 360-374. doi: 10.2307/747933

Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children’s reading skill: A five-year longitudinal study. *Child Development, 73*, 445-460.

Sénéchal, M., LeFevre, J. A., Thomas, E. M., & Daley, K. E. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly, 33*(1), 96-116.

*Sénéchal, M., Thomas, E., & Monker, J. A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology*, 218-229.

Snow, C. E., Griffin, P., & Burns, M. S. (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. San Francisco, CA: Jossey-Bass.

Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: the role of phonotactic probability and word frequency. *Journal of Child Language, 32*(04), 827-827. doi: 10.1017/S0305000905007099

*Strasser, K., Larraín, A., & Lissi, M. R. (2013). Effects of Storybook Reading Style on Comprehension: The Role of Word Elaboration and Coherence Questions. *Early Education & Development, 24*(5), 616-639. doi: 10.1080/10409289.2012.715570

*Suggate, S. P., Lenhard, W., Neudecker, E., & Schneider, W. (2013). Incidental vocabulary acquisition from stories: Second and fourth graders learn more from listening than reading. *First Language, 33*(6), 551-571. doi: 10.1177/0142723713503144

Swanborn, M. S. L., & de Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning, 52*(1), 95-117. doi: 10.1111/1467-9922.00178

*Teepe, R. C., Molenaar, I. & Verhoeven, L. (2017). Technology-enhanced storytelling stimuling parent-child interaction and preschool children's vocabulary knowledge. *Journal of Computer Assisted Learning* Advanced online publication. coi: 10.111/jcal.12169

Valdez-Menchaca, M. C., & Whitehurst, G. J. (1992). Accelerating language development through picture book reading: A systematic extension to Mexican day care. *Developmental Psychology, 28*(6), 1106-1114.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a prioir weight functions. *Psychological Methods, 10*(4), 428-443. doi: 10.1037/1082/989X.10.4.428.supp

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology, 38*(1), 246-254. doi: 10.1037/a0025260

Vuattoux, D., Japel, C., Dion, E., & Dupéré, V. (2014). Targeting the specific vocabulary needs of at-risk preschoolers: a randomized study of the effectiveness of an educator-implemented intervention. *Prevention science : the official journal of the Society for Prevention Research, 15*(2), 156-164. doi: 10.1007/s11121-013-0379-5

*Walsh, B. A., & Blewitt, P. (2006). The effect of questioning style during storybook reading on novel vocabulary acquisition of preschoolers. *Early Childhood Education Journal, 33*(4), 273-278. doi: 10.1007/s10643-005-0052-0

*Walsh, B. A., & Rose, K. K. (2013). Impact of Adult Vocabulary Noneliciting and Eliciting Questions on the Novel Vocabulary Acquisition of Preschoolers Enrolled in Head Start. *Journal of Research in Childhood Education, 27*(1), 31-45. doi: 10.1080/02568543.2012.712085

Wang, M. C., & Bushman, B. (1998). Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods, 3*(1), 46-54.

Weizman, Z. O., & Snow, C., E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*(2), 265-279. doi: 10.1037///0012-1649.37.2.265

Whitehurst, G. J., Falco, C. J., Lonigan, J. E., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology, 24*(4), 552-559. doi: 10.1037/0012-1649.24.4.552

*Wilkinson, K. S., & Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess...: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics, 34*(3), 591-613. doi: 10.1017/S0142716411000889

*Williams, S. E., & Horst, J. S. (2014). Goodnight book: sleep consolidation improves word learning via storybooks. *Frontiers in Psychology, 5*(MAR), 184-184. doi: 10.3389/fpsyg.2014.00184

*Williams, S. E., & Horst, J. S. (2015). *[Raw data and study description]*. Unpublished raw data.

*Williams, S. E., & Horst, J. S. (2015). *[Raw data and study description]*. Unpublished raw data.

Woodward, A., Markman, E., M, & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology, 30*(4), 553-566. doi: 10.1037/0012-1649.30.4.553

Zipoli, R. P., Coyne, M. D., & McCoach, D. B. (2010). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically related and embedded word review. *Remedial and Special Education, 32*(2), 131-143. doi: 10.1177/0741932510361262

Table 1. Moderators for word learning from shared storybook reading (k = 110) and with outliers removed (k = 105)

	Full Sample			Without Outliers		
Moderator Analysis	***k***	**Estimate**	**[95% BCa CIs]**	***k***	**Estimate**	**[95% BCa CIs]**
No. Target Words	110 (38)	0.062*	[0.009, 0.106]	105 (35)	0.054*	[-0.009, 0.088]
No. Words Tested	110 (38)	0.176***	[0.081, 0.537]	105 (35)	0.160***	[0.072, 0.588]
Reader						
No. target words	108 (36)	0.085**	[0.035, 0.167]	103 (33)	0.074**	[0.037, 0.124]
Reader		0.473	[-1.613, 0.604]		0.109	[-0.920, 1.070]
Dialogic Reading						
No. target words	110 (38)	0.050	[-0.001, 0.086]	105 (35)	0.036	[-0.028, 0.067]
Dialogic reading		1.224***	[0.769, 1.847]		1.207***	[0.760, 1.847]
Child's Age						
No. target words	110 (38)	0.059*	[-0.015, 0.105]	105 (35)	0.052*	[-0.012, 0.088]
Age		0.006***	[-0.001, 0.024]		0.006***	[-0.004, 0.021]
Tokens						
No. target words	83 (29)	0.329***	[-0.075, 0.559]	80 (27)	0.234**	[-0.032, 0.432]
Tokens		0.259***	[-0.037, 0.402]		0.254***	[0.026, 0.400]
Story Repetitions						
No. target words	104 (36)	0.073	[-0.042, 0.184]	101 (34)	0.036	[-0.046, 0.131]
Story repetitions		0.156***	[-0.129, 0.390]		0.162***	[-0.107, 0.363]
Story Presentation to Test Interval						
No. target words	106 (36)	0.068*	[0.009, 0.152]	101 (33)	0.046	[-0.023, 0.118]
Story to test interval		0.001	[-0.002, 0.002]		0.001**	[-0.002, 0.003]
Proportion of Nouns						
No. target words	66 (30)	0.210*	[0.097, 0.330]	63 (28)	0.131*	[0.021, 0.194]
Proportion of nouns		-2.345	[-4.548, 0.518]		-2.421*	[-4.595, -0.500]
Word Novelty						
No. target words	110 (38)	0.101***	[0.050, 0.210]	105 (35)	0.092***	[0.043, 0.166]
Novel v confirmed unknown		-1.638**	[-2.728, -0.696]		-1.120*	[-2.116, -0.463]
Novel v sophisticated		1.263*	[-0.117, 2.162]		1.654**	[0.962, 2.751]

Note: p values * < .05, ** < .01, *** < .001

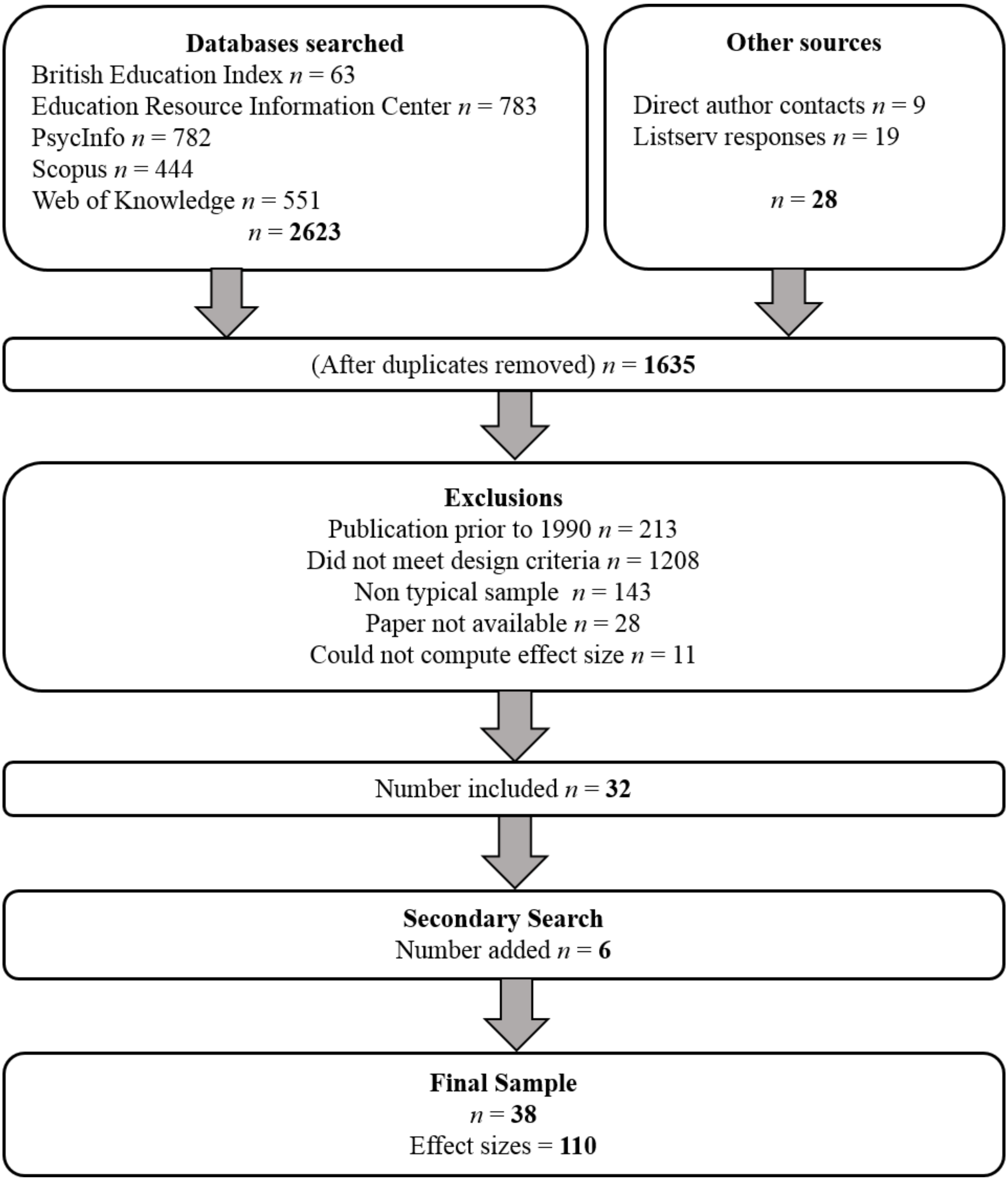

Figure 1. The schematic to show the identification process for included studies

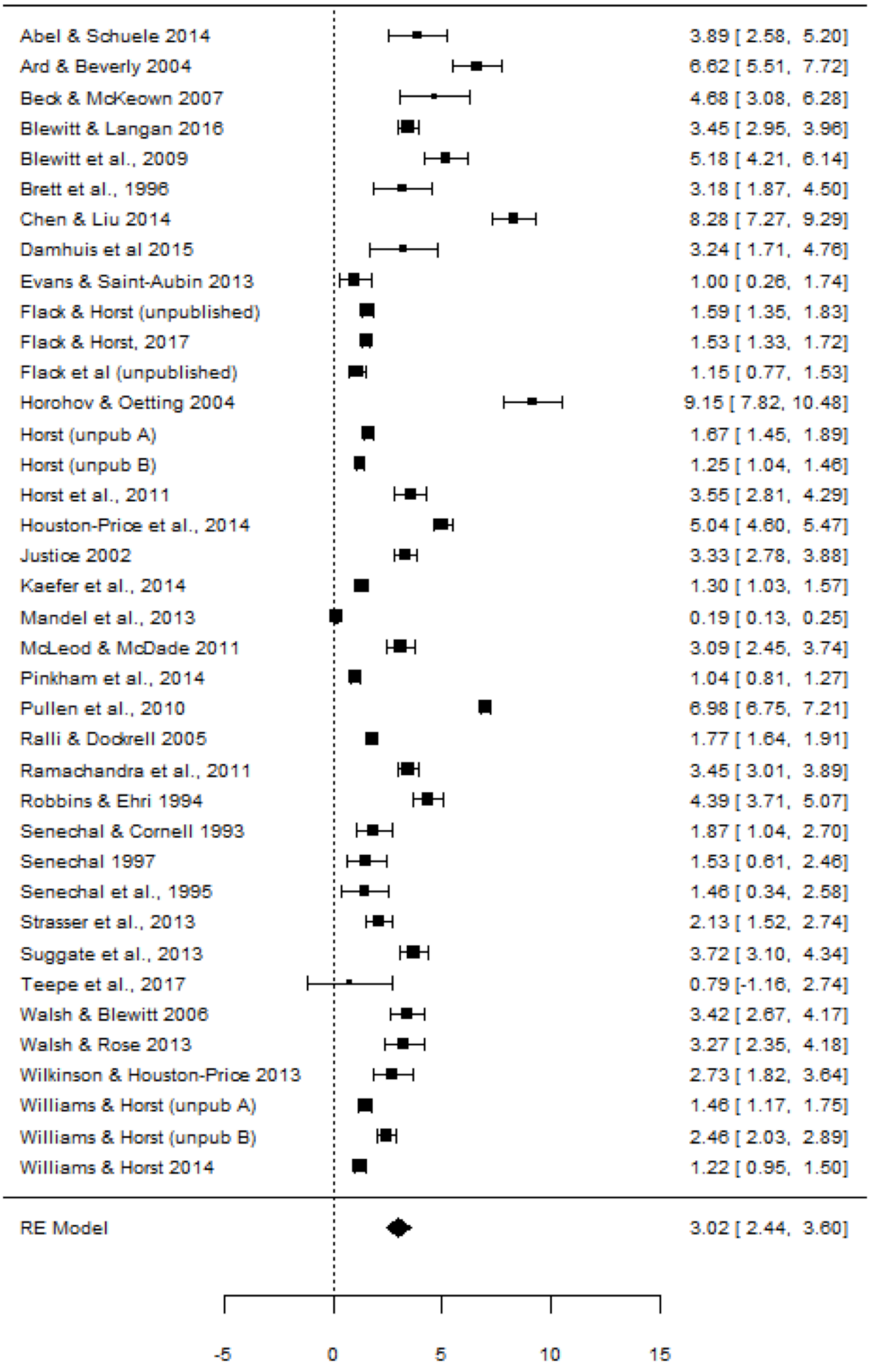

Figure 2. Effect sizes (the number of new words learned between the pre- and post-test) for the 38 studies. Note, where studies include multiple effect sizes, these are aggregated.

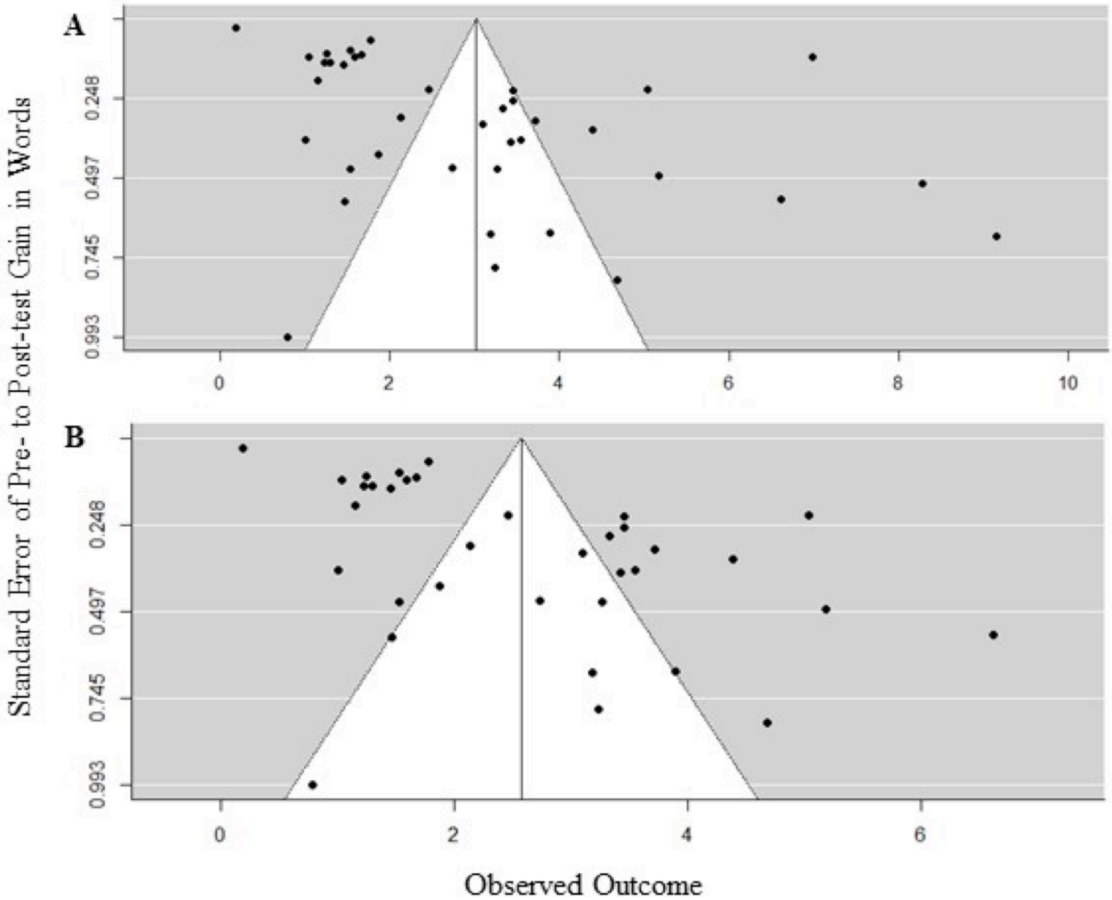

Figure 3. Funnel plots of effect sizes for the meta analysis. Panel A shows the original analysis and Panel B shows the analysis with outliers removed.